

Data Cleaning

DATABASE NAME-----USE hospital db:

Renaming/Atering The Table Column Names

ALTER TABLE facility RENAME COLUMN `Facility.Name` TO Name, RENAME COLUMN `Facility.City` TO City, RENAME COLUMN `Facility.State` TO State, RENAME COLUMN `Facility.Type` TO Type;



ALTER TABLE procedures

RENAME COLUMN `Procedure.Heart Attack.Cost` TO Heart_Attack_Cost, RENAME COLUMN `Procedure.Heart Attack.Quality` TO Heart_Attack_Quality, RENAME COLUMN `Procedure.Heart Attack.Value` TO Heart_Attack_Value, RENAME COLUMN `Procedure.Heart Failure.Cost` TO Heart_Failure_Cost, RENAME COLUMN `Procedure.Heart Failure.Quality` TO Heart_Failure_Quality, RENAME COLUMN 'Procedure. Heart Failure. Value' TO Heart Failure Value, RENAME COLUMN 'Procedure. Hip Knee. Cost' TO Hip Knee Cost, RENAME COLUMN 'Procedure. Hip Knee. Quality' TO Hip Knee Quality, RENAME COLUMN 'Procedure. Hip Knee. Value' TO Hip Knee Value, RENAME COLUMN 'Procedure.Pneumonia.Cost' TO Pneumonia Cost, RENAME COLUMN 'Procedure.Pneumonia.Quality' TO Pneumonia_Quality, RENAME COLUMN `Procedure.Pneumonia.Value` TO Pneumonia Value;

ALTER TABLE rating

RENAME COLUMN `Rating.Effectiveness` TO Effectiveness, RENAME COLUMN `Rating.Experience` TO Experience, RENAME COLUMN 'Rating.Imaging' TO Imaging, RENAME COLUMN `Rating.Mortality` TO Mortality, RENAME COLUMN `Rating.Overall` TO Overall, RENAME COLUMN 'Rating.Readmission' TO Readmission, RENAME COLUMN 'Rating.Safety' TO Safety, RENAME COLUMN `Rating.Timeliness` TO Timeliness;



Update Invalid Ratings (e.g., -1)

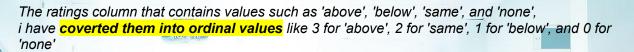
UPDATE rating SET rating.Overall = NULL WHERE rating.Overall = - 1;

#Remove Rows with Zero Procedure Costs

DELETE FROM procedures WHERE procedures. Heart Attack Cost = 0 OR procedures. Heart Failure Cost = 0 OR procedures.Hip_Knee_Cost = 0 OR procedures.Pneumonia_Cost = 0;

#Remove Rows with Zero Ratings

DELETE FROM rating
WHERE
Overall = 0
OR Mortality = 0
OR Safety = 0
OR Readmission = 0
OR Experience = 0
OR Effectiveness = 0
OR Timeliness = 0
OR Imaging = 0;



foe example:-

UPDATE rating
SET
rating.Mortality = 0
WHERE
rating.Mortality = 'None';

UPDATE rating
SET
rating.Mortality = 3
WHERE
rating.Mortality = 'above';

Data Exploration

DESCRIBE facility;

Re	sult Grid	Filter F	Rows:			Export:		Wra
	Field	Type	Null	Key	Default	Extra		
•	Name	text	YES		NULL			
	City	text	YES		NULL			
	State	text	YES		NULL			
	Туре	text	YES		NULL			
	Facility_ID	int	NO	PRI	NULL	auto_ind	remer	nt

DESCRIBE procedures;

_	1 1 EE			CONTRACTOR OF THE PERSON NAMED IN COLUMN 1		H
Re	sult Grid 🔢 Filter Rov	/S:			Export: #	Wrap
	Field	Type	Null	Key	Default	Extra
•	Heart_Attack_Cost	int	YES		NULL	
	Heart_Attack_Quality	text	YES		NULL	
	Heart_Attack_Value	text	YES		NULL	
	Heart_Failure_Cost	int	YES		NULL	
	Heart_Failure_Quality	text	YES		NULL	
	Heart_Failure_Value	text	YES		NULL	
	Pneumonia_Cost	int	YES		NULL	
	Pneumonia_Quality	text	YES		NULL	
	Pneumonia_Value	text	YES		NULL	
	Hip_Knee_Cost	int	YES		NULL	
	Hip_Knee_Quality	text	YES		NULL	
	Hip_Knee_Value	text	YES		NULL	
	Procedure_ID	int	NO	PRI	NULL	auto
	overall_procedure_a	double	YES		NULL	
	Combined_Quality	text	YES		NULL	
	Facility_ID	int	YES	MUL	NULL	

DESCRIBE rating;

Re	sult Grid 🔠 🛚	Filter Rov	/s:		Б	oport: 📳 Wrap (
	Field	Type	Null	Key	Default	Extra
•	Overall	int	YES		NULL	
	Mortality	text	YES		NULL	
	Safety	text	YES		NULL	
	Readmission	text	YES		NULL	
	Experience	text	YES		NULL	
	Effectiveness	text	YES		NULL	
	Timeliness	text	YES		NULL	
	Imaging	text	YES		NULL	
	Rating_ID	int	NO	PRI	NULL	auto_increment
	Facility_ID	int	YES	MUL	NULL	

These queries provide the structure of the facility, procedures, and rating tables, listing columns, their data types, and constraints.

#Count Total Rows and Distinct Facilities

SELECT
COUNT(*) AS Total_Rows,
COUNT(DISTINCT facility.Name) AS Distinct_Facility
FROM
facility;

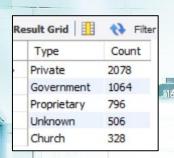
	Total_Rows	Distinct_Facility
•	4772	4589

This query counts the total number of rows in the facility table and the number of distinct facilities by name.

#Check Distribution of Facility Types

SELECT

facility.Type, COUNT(*) AS Count
FROM
facility
GROUP BY facility.Type
ORDER BY count DESC;



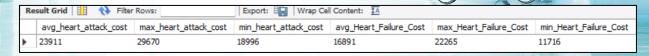
This query groups the facility table by Type and counts the number of facilities for each type, ordered from most to least frequent.

Summary Statistics for All Procedure Costs

SELECT

ROUND(AVG(procedures.Heart_Attack_Cost)) AS avg_heart_attack_cost, MAX(procedures.Heart_Attack_Cost) AS max_heart_attack_cost, MIN(procedures.Heart_Attack_Cost) AS min_heart_attack_cost, ROUND(AVG(procedures.Heart_Failure_Cost)) AS avg_Heart_Failure_Cost, MAX(procedures.Heart_Failure_Cost) AS max_Heart_Failure_Cost, MIN(procedures.Heart_Failure_Cost) AS min_Heart_Failure_Cost FROM

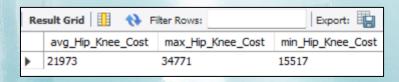
procedures;



SELECT

ROUND(AVG(procedures.Hip_Knee_Cost)) AS avg_Hip_Knee_Cost, MAX(procedures.Hip_Knee_Cost) AS max_Hip_Knee_Cost, MIN(procedures.Hip_Knee_Cost) AS min_Hip_Knee_Cost FROM

procedures;

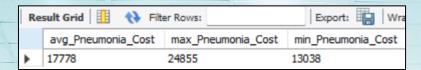


SELECT

ROUND(AVG(procedures.Pneumonia_Cost)) AS avg_Pneumonia_Cost, MAX(procedures.Pneumonia_Cost) AS max_Pneumonia_Cost, MIN(procedures.Pneumonia_Cost) AS min_Pneumonia_Cost

FROM

Procedures;



This query calculates the average, maximum, and minimum costs for heart attack,heart failure procedures,Hip Knee & Pneumonia.

#Count of Ratings

SELECT
Overall, COUNT(*) AS count
FROM
rating
GROUP BY Overall
ORDER BY Overall;

Result Grid					
	Overall	count			
٠	1	237			
	2	598			
	3	737			
	4	588			
	5	205			

Data Integration

#Create a New Column for Average Procedure Cost

alter table procedures
add column overall_procedure_avg_cost real;

UPDATE procedures

overall_procedure_avg_cost = ((procedures.Heart_Attack_Cost + procedures.Heart_Failure_Cost

+ procedures.Hip_Knee_Cost + procedures.Pneumonia_Cost) / 4);

overall_procedure_avg_cost
21132
19441
19189.75
20549.5
20001.5
20428
20792.5

Combine Quality Ratings into a Single Column

ALTER TABLE procedures
ADD COLUMN Combined_Quality TEXT;

UPDATE procedures
SET Combined_Quality = concat(
 procedures.Heart_Attack_Quality,', ',
 procedures.Heart_Failure_Quality,', ',
 procedures.Hip_Knee_Quality,', ',
 procedures.Pneumonia_Quality);

Combined_Quality

Average, Average, Average, Average Average, Average, Average, Average Average, Worse, Average, Worse Average, Average, Average Average, Average, Average Average, Average, Average Average, Average, Average Average, Average, Average

Statistical Analysis

#Correlation Between Overall Rating and Procedure Costs

SELECT

rating.Overall,

AVG(procedures.Heart_Attack_Cost) AS avg_Heart_Attack_Cost, AVG(procedures.Heart_Failure_Cost) AS avg_Heart_Failure_Cost, AVG(procedures.Hip_Knee_Cost) AS avg_Hip_Knee_Cost, AVG(procedures.Pneumonia_Cost) AS avg_Pneumonia_Cost

FROM

procedures

JOIN

rating ON procedures.Facility_ID = rating.Facility_ID GROUP BY rating.Overall ORDER BY rating.Overall;

Re	sult Grid	Filter Rows:	Export:	Wrap Cell Content:	<u>‡A</u>
	Overall	avg_Heart_Attack_Cost	avg_Heart_Failure_Cost	avg_Hip_Knee_Cost	avg_Pneumonia_Cost
•	1	24111.2245	17075.1020	21975.5102	180 18.3673
	2	23899.8049	16968.0244	22020.3310	17734.4634
	3	23908.0097	16821.6052	21958.0971	17824.4239
	4	23914.7424	16831.7614	22198.4091	17823.9091
	5	23616.2750	16630.2500	21925.0875	17730.1625

#Distribution of Ratings by Facility Type

SELECT

facility.Type, rating.Overall, COUNT(*) AS count FROM

facility

JOIN

rating ON facility.Facility_ID = rating.Facility_ID GROUP BY facility.Type , rating.Overall ORDER BY facility.Type , rating.Overall;

Re	sult Grid	Filter	r Rows:
	Туре	Overall	count
•	Church	1	15
	Church	2	37
	Church	3	45
	Church	4	47
	Church	5	16
	Government	1	43
	Government	2	129
	Government	3	170
	Government	4	137
	Government	5	46
	Private	1	108
	Private	2	275
	Private	3	323
	Private	4	247
	Private	5	83
	Proprietary	1	43
	Proprietary	2	99



Heart Attack Quality Distribution

SELECT

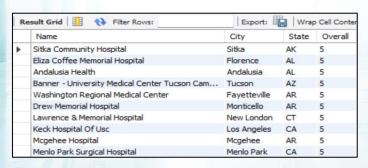
Heart_Attack_Quality, COUNT(*) AS count
FROM

procedures
GROUP BY Heart_Attack_Quality
ORDER BY count DESC;

Re	sult Grid 🔢 🙌 Filt	er Rows:
	Heart_Attack_Quality	count
•	Average	1947
	Better	43
	Worse	23
	Unknown	2

#Top 10 Hospitals by Overall Rating

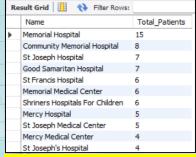
SELECT
f.Name, f.City, f.State, r.Overall
FROM
facility f
JOIN
rating r ON f.Facility_ID = r.Facility_ID
ORDER BY r.Overall DESC
LIMIT 10;



Track Patient Records

#Total Number of Patients Per Facility

SELECT
Name, COUNT(*) AS Total_Patients
FROM
facility
GROUP BY Name
ORDER BY Total_patients DESC;



#Patient Distribution by State:

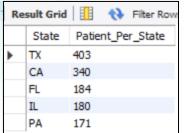
SELECT

State, COUNT(*) AS Patient_Per_State FROM

facility

GROUP BY State

ORDER BY Patient_Per_State DESC LIMIT 5;



#HERE TX STAND FOR TEXAS WITH 403 TOTAL PATIENTS

Evaluating Healthcare Metrics

#Average Rating of Facilities:

SELECT

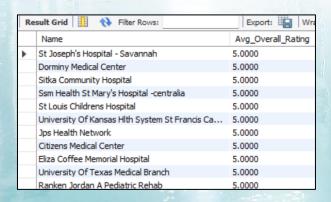
facility.Name, AVG(rating.Overall) AS Avg_Overall_Rating FROM

facility

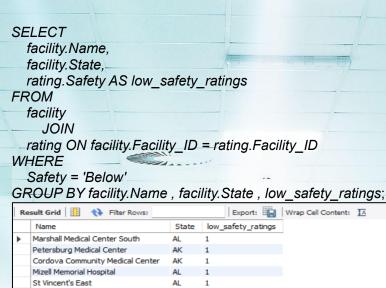
JOIN

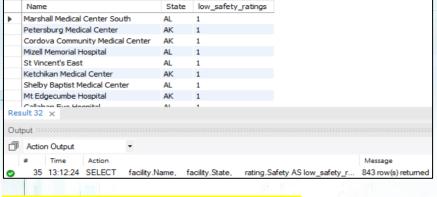
rating ON facility.Facility_ID = rating.Facility_ID GROUP BY facility.Name

ORDER BY Avg_Overall_Rating DESC;



#Identify Facilities with Low Safety Ratings:





#Calculate Average Cost of Each Procedures:

SELECT

f.Name, AVG(p.Heart_Attack_Cost) AS avg_heart_attack_cost FROM

facility f

JOIN

procedures p ON f.Facility_ID = p.Facility_ID GROUP BY f.Name ORDER BY avg_heart_attack_cost DESC;



SELECT

f.Name, AVG(p.Heart_Failure_Cost) AS avg_Heart_Failure_Cost FROM

facility f

JOIN

procedures p ON f.Facility_ID = p.Facility_ID

GROUP BY f.Name

ORDER BY avg Heart Failure Cost DESC;



SELECT

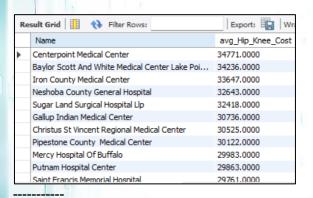
f.Name, AVG(p.Hip_Knee_Cost) AS avg_Hip_Knee_Cost FROM

facility f

JOIN

procedures p ON f.Facility_ID = p.Facility_ID GROUP BY f.Name

ORDER BY avg_Hip_Knee_Cost DESC;



SELECT

f.Name, AVG(p.Pneumonia_Cost) AS avg_Pneumonia_Cost FROM

facility f

JOIN

procedures p ON f.Facility_ID = p.Facility_ID GROUP BY f.Name

ORDER BY avg_Pneumonia_Cost DESC;

	Re	sult Grid H N Filter Rows:	Export: Wrap
		Name	avg_Pneumonia_Cost
	•	Baylor Scott And White Medical Center Lake Poi	24855.0000
		Seneca District Hospital	24200.0000
		Baylor Medical Center At Frisco	23983.0000
3		Jupiter Medical Center	23778.0000
		Indianhead Medical Center Shell Lake	23258.0000
		Palm Bay Hospital	23252.0000
		Mercy Medical Center Mt Shasta	23192.0000
		Fairchild Medical Center	23030.0000
		John T Mather Memorial Hospital Of Port Jeffer	22973.0000
		Bon Secours Depaul Medical Center	22909.0000
l		Upland Hills Health	22872.0000

Identify Areas for Improvement

#Facilities with High Readmission Rates:

SELECT DISTINCT

f.Name, f.State, r.Readmission, r.Overall AS Overall_Rating FROM

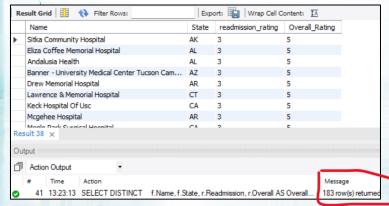
facility f

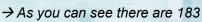
JOIN

rating r ON f.Facility_ID = r.Facility_ID

WHERE

Readmission = 'Above' AND r.Overall = 5 ORDER BY Overall Rating DESC;





rows with the oveall rating of 5 & readmilision rating 3 on the scale of (1-3)

#Patient Distribution by State(TOP 5):

SELECT

f.State, COUNT(f.State) AS Patient_from_each_state

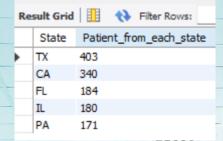
FROM

facility f

GROUP BY f.State

ORDER BY Patient_from_each_state DESC

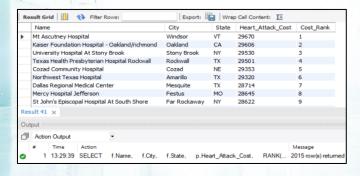
LIMIT 5;



Ranking Facilities by Procedure Cost

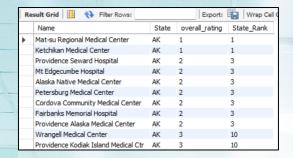
#Rank Facilities by Heart Attack Procedure Cost

```
SELECT
f.Name,
f.City,
f.State,
p.Heart_Attack_Cost,
RANK() OVER (ORDER BY p.Heart_Attack_Cost DESC) AS Cost_Rank
FROM
facility f
JOIN
procedures p
ON f.Facility_ID = p.Facility_ID
ORDER BY
Cost_Rank;
```



#Ranking Facilities by Overall Rating within Each State

```
SELECT
f.Name,
f.State,
r.Overall AS overall_rating,
RANK() OVER (PARTITION BY f.State ORDER BY r.Overall) AS State_Rank
FROM
facility f
JOIN
rating r
ON f.Facility_ID = r.Facility_ID
ORDER BY
f.State,
State_Rank;
```



Calculating Running Totals and Averages

#Running Total of Heart Attack Procedure Costs

```
SELECT
  f.Name,
  f.City,
  f.State,
  p.Heart_Attack_Cost,
  SUM(p.Heart_Attack_Cost) OVER (ORDER BY p.Heart_Attack_Cost) AS Running_total_cost
FROM
  facility f
JOIN
```

procedures p ON f.Facility_ID = p.Facility_ID

ORDER BY Running_total_cost;

Export: Wrap Cell Content: TA State Heart_Attack_Cost Running_total_cost MO 19128 Moberly Regional Medical Center Moberly 38124 Seton Edgar B Davis Hospital 57707 Luling TX 19583 University Of Maryland Medical Center Baltimore Russell County Hospital Russell Springs 19680 97014

LA 116886 Pointe Coupee General Hospital New Roads 19872 136833 Helen Hayes Hospital West Haverstraw Sanford ND 19991 156824 Muleshoe Area Medical Center Muleshoe TX 20088 176912 Sanford Bagley Medical Center Bagley MN 20092 197004

#Moving Average of Heart Failure Procedure Cost

```
SELECT
  f.Name,
  f. City,
  f.State.
  p.Heart Failure Cost,
  AVG(p.Heart_Failure_Cost) OVER (ORDER BY p.Heart_Failure_Cost) AS Moving_Avg_cost
FROM
  facility f
JOIN
  procedures p
  ON f.Facility_ID = p.Facility_ID
ORDER BY
  Moving_Avg_cost;
```

ſ	Re	sult Grid 🎚 🛟 Filter Rows:	Export	: 📳 V	Wrap Cell Content: TA	
		Name	City	State	Heart_Failure_Cost	Moving_Avg_cost
I	•	Seton Edgar B Davis Hospital	Luling	TX	11716	11716.0000
		Kings Daughters Medical Center Ohio	Portsmouth	OH	12947	12331.5000
		Maimonides Medical Center	Brooklyn	NY	13054	12572.3333
		Scripps Memorial Hospital - Encinitas	Encinitas	CA	13122	12709.7500
		St. Joseph's Hospital And Medical Center	Phoenix	AZ	13287	12825.2000
		Glens Falls Hospital	Glens Falls	NY	13374	12916.6667
		University Hospital Of Brooklyn (Downstate)	Brooklyn	NY	13391	12984.4286
4		Muleshoe Area Medical Center	Muleshoe	TX	13420	13038.8750
		John C. Lincoln North Mountain Hospital	Phoenix	AZ	13788	13122.1111
		Mercy Medical Center-clinton	Clinton	IA	13811	13191.0000
L		Wood County Hospital	Bowling Gr	OH	13830	13249.0909

Cumulative Metrics

#Cumulative Count of Facilities by Rating

SELECT f.Name,

f. City,

f.State,

r.Overall,

COUNT(*) OVER (ORDER BY r.Overall) AS cumulative_count

facility f

JOIN

rating r

ON f.Facility_ID = r.Facility_ID

ORDER BY

cumulative_count;

Name	City	State	Overall	cumulative_count
Texas General Hospital- Vzrmc Lp	Grand Saline	TX	1	237
Providence Hospital Of North Houst	Houston	TX	1	237
The Hospitals Of Providence Transm	El Paso	TX	1	237
Saint Camillus Medical Center	Hurst	TX	1	237
Providence Seward Hospital	Seward	AK	2	835
Marshall Medical Center South	Boaz	AL	2	835
Petersburg Medical Center	Petersburg	AK	2	835

#Cumulative Average of Procedure Costs by Facility Type

SELECT

f. Type,

f.Name,

p.Pneumonia_Cost,

AVG(p.Pneumonia_Cost) OVER (PARTITION BY f.Type ORDER BY f.Name) AS

cumulative_avg_cost

FROM

facility f

JOIN

procedures p

ON f.Facility_ID = p.Facility_ID

ORDER BY

f.Type,

f.Name;

R	esult Grid	Filter Rows:	Export: Wrap Ce	ell Content: IA	
	Туре	Name	Pneumonia_Cost	cumulative_avg_cost	
•	Church Adventist Health Castle		18884	18884.0000	
	Church	Adventist Health Feather River	18842	18863.0000	
	Church	Adventist Health Hanford	17705	18477.0000	
	Church	Adventist Health Howard Memorial	18594	18506.2500	
	Church	Adventist Health Simi Valley	18358	18476.6000	
	Church	Adventist Health Tillamook	17348	18288.5000	
	Church	Adventist La Grange Memorial Hospital	17444	18167.8571	

Identifying Top and Bottom Performers

#Top 5 Facilities by Hip/Knee Procedure Cost within Each City

```
WITH RankedFacilities AS (
                                                       SELECT
    f.Name,
    f. City,
    f.State,
    p.Hip_Knee_Cost, 165 171
    RANK() OVER (PARTITION BY f.City ORDER BY p.Hip_Knee_Cost DESC) AS
City_cost_rank
  FROM
    procedures p
  JOIN
    facility f ON p.Facility_ID = f.Facility_ID
SELECT
  Name,
  City,
  State,
  Hip_Knee_Cost,
  City_cost_rank
FROM
```

RankedFacilities

WHERE

City_cost_rank <= 5

ORDER BY

City, City_cost_rank limit 5;

Result Grid Filter Rows:			Export: Wrap Cell Content: TA			
	Name	City	State	Hip_Knee_Cost	City_cost_rank	
٨	Abbeville General Hospital	Abbeville	LA	20180	1	
	Grays Harbor Community Hospital	Aberdeen	WA	22202	1	
	Avera St Lukes	Aberdeen	SD	19894	2	
	Memorial Hospital	Abilene	KS	23497	1	
	Abilene Regional Medical Center	Abilene	TX	20914	2	

#Bottom 5 Facilities by Overall Rating

```
WITH RankedFacilities AS (
```

SELECT

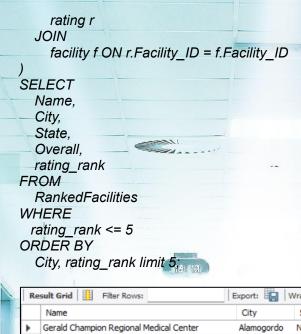
f.Name,

f.City,

f.State,

r.Overall,

RANK() OVER (ORDER BY r.Overall asc) AS rating_rank FROM



Result Grid Filter Rows: Export: Rows: Wrap Cell Content:							
	Name	City	State	Overall	rating_rank		
•	Gerald Champion Regional Medical Center	Alamogordo	NM	1	1		
	Mayo Clinic Health System - Albert Lea And Austin	Albert Lea	MN	1	1		
	Central Louisiana Surgical Hospital	Alexandria	LA	1	1		
	Island Hospital	Anacortes	WA	1	1		
	Hospital Dr Susoni Inc	Arecibo	PR	1	1		
		118			171		